Reward Crowdfunding Success Forecasting: An Empirical Evaluation of Machine Learning Algorithms

Autoria

Wesley Mendes-da-Silva - mr.mendesdasilva@gmail.com
Mestr e Dout em Admin de Empresas / FGV/EAESP - Fundação Getulio Vargas - Escola de Administração de Empresas de São Paulo

Israel José dos Santos Felipe - israeljfelipe@gmail.com
Prog de Pós-Grad em Admin – PPGA / UFRN - Universidade Federal do Rio Grande do Norte
64 / NIPE/Universidade do Minho (Portugal)

George Darmiton da Cunha Cavalcanti - gdcc@cin.ufpe.br
Programa de Pós-graduação em Ciência da Computação / UFPE - Universidade Federal de Pernambuco

Leonardo Alves dos Santos - las3@cin.ufpe.br
Programa de Pós-Graduação em Ciência da Computação / UFPE - Universidade Federal de Pernambuco

Agradecimentos

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Resumo

Reward-based crowdfunding has been increasingly used by entrepreneurs and small businesses to raise capital for their creative projects, whose success is central to this industry. We offer new empirical evidence from machine learning (ML) models, based on variables related to sentiment on media campaign profile, and geography of the cities where the campaigns are launched. We evaluated fifteen ML models and in particular multiple classifier systems (MCSs), a promising category of algorithms that combine different models. Over 4,000 campaigns hosted on the largest reward crowdfunding platform in one of the ten largest economies in the world are used in the experimental study. Our results show that Meta-DES, which performs dynamic selection, obtains the best overall results among the evaluated models such as artificial neural networks, decision trees, random forest and XGBoost. Furthermore, since usually interpreting the output of ML models is considered to be very difficult due to their complex “black box” architecture, we also use Shapley additive explanations to interpret the outputs of forecasting. Among variables evaluated in our models, including the textual sentiment of the mass media, the number of pledges and the target amount of the campaign deserve to be highlighted when predicting the campaign’s success.
ARTICLE TEMPLATE

Reward Crowdfunding Success Forecasting: An Empirical Evaluation of Machine Learning Algorithms

ARTICLE HISTORY
Compiled April 14, 2022

ABSTRACT
Reward-based crowdfunding has been increasingly used by entrepreneurs and small businesses to raise capital for their creative projects, whose success is central to this industry. We offer new empirical evidence from machine learning (ML) models, based on variables related to sentiment on media campaign profile, and geography of the cities where the campaigns are launched. We evaluated fifteen ML models and in particular multiple classifier systems (MCSs), a promising category of algorithms that combine different models. Over 4,000 campaigns hosted on the largest reward crowdfunding platform in one of the ten largest economies in the world are used in the experimental study. Our results show that Meta-DES, which performs dynamic selection, obtains the best overall results among the evaluated models such as artificial neural networks, decision trees, random forest and XGBoost. Furthermore, since usually interpreting the output of ML models is considered to be very difficult due to their complex “black box” architecture, we also use Shapley additive explanations to interpret the outputs of forecasting. Among variables evaluated in our models, including the textual sentiment of the mass media, the number of pledges and the target amount of the campaign deserve to be highlighted when predicting the campaign’s success.

KEYWORDS
Crowdfunding; Machine Learning; Artificial Intelligence; Ensemble Learning; Explainable AI; Forecasting.
1. Introduction

Crowdfunding has occupied a growing space in the literature of several fields of knowledge, from finance (Sorenson, Assenova, Li, Boada, & Fleming, 2016) to art (Maechling, 2018) and information systems (C. Bogusz, Laurell, Sandström, & Sandström, 2020; C. I. Bogusz, Teigland, & Vaast, 2019; Sutanto, Wenninger, & Duriana, 2021) thus demonstrating society’s interest in this modality of fast and flexible financing (Felipe, Mendes-Da-Silva, Leal, & Braun Santos, 2022; Luo, Ge, & Wang, 2022) around the world (Burtch, Ghose, & Wattal, 2014). Therefore, predicting the success of crowdfunding campaigns is a topic to which researchers around the world have dedicated efforts (Kim, Kannan, Trusov, & Ordanini, 2020; Mendes-Da-Silva, Rossoni, Conte, Gattaz, & Francisco, 2016; Yan et al., 2018). Historically, the prediction of success of crowdfunding reward campaigns has focused on traditional methods, such as linear regression or logit, which are largely dependent on parameterization and a set of assumptions. As a consequence, model results may end up being not very robust (Babin, Ortinau, Herrmann, & Lopez, 2021; Ethiraj, Gambardella, & Helfat, 2016; Honig & Samuelsson, 2015; Tsang & Kwan, 1999). In parallel, machine learning (ML) techniques have been pointed out as a means to obtain more accurate predictions for the success of campaigns. However, research on this topic is still virtually rare (Peng, Zhou, Niu, & Feng, 2021; Shafqat, Byun, & Park, 2020; J.-Y. Yeh & Chen, 2020; H. Yuan, Lau, & Xu, 2016).

Many ML models are available, and choosing one that best fits a given problem is challenging. Thus, instead of adopting only one model, Multiple Classifier System (MCS) (L. I. Kuncheva, 2014) employs many models to mitigate the uncertainties related to choosing a single model. The fundamental idea is combining the performance of different models to improve the overall accuracy. MCS has proven its ability to obtain better precision in comparison to single machine learning models (Woźniak, Graña, & Corchado, 2014a). It can be divided into static and dynamic types. The former combines all the models to predict a test instance, while the latter dynamically selects a subset of the best models per test instance to perform the prediction. Among the MCSs, dynamic selection has obtained promising results (Britto, Sabourin, & Oliveira,
This paper assesses reward crowdfunding campaign success using Machine Learning models, particularly multiple classifier systems that dynamically select the best subset of models for each campaign under analysis. Thus, we compare the performance of single models with static and dynamic multiple classifiers systems in the context of predicting crowdfunding campaign success.

Crowdfunding has become an increasingly more important channel for entrepreneurs to raise funds from the crowd to support their startup projects (Böckel, Hörisch, & Tenner, 2021; Mochkabadi & Volkmann, 2020; Wehnert & Beckmann, 2021). Previous studies examined various factors, such as project features, geographic aspects and categories of projects that might influence the outcomes of the fund raising campaigns (Kaminski & Hopp, 2020; Yin, Liu, & Mirkovski, 2019). However, textual information of projects has rarely been studied for analyzing crowdfunding successes (W. Wang, Zhu, Wang, & Wu, 2017; H. Yuan et al., 2016). Furthermore, the study of textual campaign information is rarely considered, on the other hand textual information in the mass media is virtually absent from the literature. This topic is of interest not only to researchers, but also to regulators and other stakeholders in the crowdfunding industry and its role (Carvajal, García-Avilés, & González, 2012; Chen, Dai, Yao, & Li, 2019; Stiver, Barroca, Minocha, Richards, & Roberts, 2015).

Researchers are interested in new methods and new variables that predict campaign success (Stasik, Wilczyńska, et al., 2018). In addition to employing ML and SHaply Additive exPlanation (SHAP), to promote interpretability of the results from an ML approach usually viewed as a 'black box' (Slack, Hilgard, Jia, Singh, & Lakkaraju, 2020), our study is a pioneering one in that it considers market sentiment reflected in mass media. Entrepreneurs and platform managers can learn from our results, as we bring new results about predicting campaign success. Regulators and policy-makers can benefit from our results for the design of public policies that can promote and disseminate the use of crowdfunding reward platforms as a fast and flexible financing channel within the space of FinTechs, especially in the economies of developing countries (Cumming & Zhang, 2016; Dorfleitner, Oswald, & Röhe, 2020; Knewton & Rosenbaum, 2020). Furthermore, crowdfunding can be a valuable tool to promote the
resilience of communities by allowing quick access to the capital needed to react in
times of crisis, such as pandemics and catastrophic events, whether caused by man or
natural events (Behl & Dutta, 2020; Lemos & Ramos, 2016; Trovato, 2021).

We perform a comprehensive set of experiments on the Brazilian reward crowdfund-
ing platform Catarse dataset composed of 4,193 campaigns. Fifteen machine learning
models are evaluated using three different metrics: Accuracy, Area Under the Receiver
Operating Characteristic (ROC) Curve, and F1-Score. Additionally, we show that the
results obtained by the Meta-DES model (R. M. O. Cruz, Sabourin, Cavalcanti, &
Ren, 2015), which is a framework that employs a dynamic selection of the models per
test instance, compare favorably to other models. After analyzing the influence of vari-
ables using SHAP, a method to explain the model’s predictions, we have observed that
campaigns that attract fewer pledges and set higher revenue targets tend to have more
difficulty obtaining success. These two variables turn out to be the main predictors of
campaign success, although there are other relevant variables such as the city’s gross
domestic product per capita, population, and also media sentiment on the campaign
launch day.

We make several contributions to the literature. The main ones are the following.
First, our research work is the design of a novel text analytic-based framework that
can extract latent semantics from the texts in mainstream and social media news.
More specifically, we develop the Domain-Constraint Latent Dirichlet allocation (DC-
LDA) topic model for effective extraction of topical features from texts of mainstream
and social media as a proxy for market sentiment. Second, we make a comprehensive
evaluation of different machine learning models in the context of reward crowdfunding
campaign success prediction. In particular, the assessment of multiple classifier systems
that perform dynamic selection as being the best classifiers in the ensemble. Third,
we carry out our study in the context of one of the main emerging economy coun-
tries. Such countries are often underrepresented in the literature (Yapprak & Karademir,
2010). Forth, we carry out an analysis to better understand how the machine learning
model’s predictions behave. This is, of great interest since it helps to interpret what
are the influence of the attributes in the campaign success. Fifth, while J.-Y. Yeh and
Chen (2020) propose a static combination of neural networks to predict the success
of crowdfunding and show that their system outperforms three monolithic classifiers, namely, linear regression, logistic regression, and artificial neural networks, on the other hand we go further and explore a literature set of methods (called dynamic selection methods) and show that they attain better accuracy rates than static combination methods.

The rest of this paper is structured as follows: Section 2 presents the fundamentals and related literature, Section 3 describes the procedure we use to develop our research work. In Section 4 we present our results, including SHAP-based explanations to understand the main results for the model with the better fit. We also summarize the results for a variety of models based on different ML procedures. Finally, Section 5 presents our concluding remarks, with suggestions for future research.

2. Related literature

2.1. Reward crowdfunding campaign success forecasting

In accordance with the theoretical fundamentals of crowdfunding (Strausz, 2017), the prediction of reward crowdfunding campaign success is a key to the development of the crowdfunding industry. This is because the prediction of success can assist individuals and organizations not only in their decision-making regarding the allocation of campaign resources as also in the recommending more successful campaigns for individual supporters. According to Strausz (2017), campaign success can be defined as follows. The entrepreneur is first asked to describe the following three elements of his or her campaign on the platform’s public webpage: i) a description of the reward to the consumer, which is typically the entrepreneur’s final product; ii) a pledge level $p$; and iii) a target amount $T_A$.

After describing these elements, the crowdfunding campaign starts and, for a fixed period of time, a consumer (backer or pledger) can pledge an amount $p$ to support the campaign financially. During the campaign, the platform provides accurate information on the current aggregate level of pledges so that a consumer can, in principle, condition his or her decision to pledge on the contributions of previous consumers. After the campaign ends, the platform compares the target amount $T_A$ to the sum of pledges
\[ P \equiv n \cdot p, \] where \( n \) is the number of pledging consumers (backers). If aggregate pledges \( P \) fall short of target level \( T_A \), the platform declares the crowdfunding campaign a failure.

To the best of our knowledge, the literature (Table 1) has not taken into consideration the determinants of campaigns to attain amount \( P > T_A \) based on Machine Learning procedure, which is the main reason that this study is considered relevant, since it contributes to the development and consolidation of the theory of reward crowdfunding success. In estimating the success of campaigns, there are two particularly relevant aspects: the classes of the explanatory variables adopted in the models and the classes of the models employed to estimate success.

\[ \text{TABLE 1 ABOUT HERE} \]

In terms of the variables, the literature points to various levels of analysis, ranging from the country level (Glazer & Konrad, 1996) to the individual level (Crosetto & Regner, 2018; Fisk et al., 2011; Hu, Li, & Shi, 2015). Here we address campaign characteristics and the cities where they are located to test the ability of machine learning models to predict campaign success (Yan et al., 2018).

2.2. Machine learning algorithms

Machine learning is a class of algorithms that extracts information from data. Typical algorithms are decision trees, multilayer perceptrons, support vector machines, k-nearest neighbor and random forests (Fernández-Delgado, Cernadas, Barro, & Amorim, 2014). These data-driven algorithms map the input to the desired prediction, and commonly only one algorithm (monolithic classifier) is applied to the given task. However, MCS, which combines many learning algorithms, has been proven to outperform monolithic classifiers in a myriad of domains (R. M. Cruz et al., 2018; Dietterich, 2000; L. Kuncheva, 2002; Polikar, 2006).

MCS has three phases (Britto et al., 2014): i) generation: the training set is used to generate a pool of classifiers; ii) selection: a subset of the pool of classifiers is selected to
perform the classification; ii) combination/integration: the predictions of the selected classifiers are combined to output the final decision.

Concerning the selection phase, it can be static or dynamic. In the static approach, no selection is performed to obtain the generalization; in other words, all the classifiers defined during the training process are combined to predict the label of the query instance. In contrast, the dynamic selection approach aimed at finding the best subset of classifiers per query instance, and such a procedure is randomly performed.

Dynamic selection (DS) can be divided into dynamic classifier selection (DCS) and Dynamic Ensemble Selection (DES). DCS selects only one classifier per new test sample, and DES selects one or more classifiers for classifying each new test sample. DS deserves special attention since it has achieved superior performance compared with monolithic classifiers and traditional static combination approaches (R. M. Cruz et al., 2018; R. M. O. Cruz et al., 2015; Ko, Sabourin, & Britto, Jr., 2008; Souza, Cavalcanti, Cruz, & Sabourin, 2019).

Meta-DES (R. M. Cruz, Sabourin, & Cavalcanti, 2017; R. M. O. Cruz et al., 2015) is a DES framework that achieves state-of-the-art performance. This framework is based on meta-learning, and its main contribution is to use a meta-classifier to assess and select the best subset of the classifier pool instead of using static selection rules as in other algorithms, such as overall local accuracy (OLA) (Woods, Kegelmeyer, & Bowyer, 1997) and k-nearest Oracle (KNORA) (Ko et al., 2008).

2.3. Machine learning explainability

Explainability is a major concern in the machine learning research field since it aims to unfold the often called black-box ML models. Methods such as Local Interpretable Model-agnostic Explanation (LIME) and SHaply Additive exPlanation (SHAP) are used to assess the variable’s contribution to the prediction. Given that local interpretations using LIME are prone to sampling variability and instability (S. Lundberg, Erion, Chen, & et al., 2020), we herein use SHAP (Bloch & Friedrich, 2021; S. M. Lundberg & Lee, 2017; Vriens, Vidden, & Bosch, 2021).

SHAP is a method to explain individual predictions that is based on the game theoretically optimal Shapley values (Shapley, 2016). The SHAP’s goal is to explain
the prediction of an instance \( x \) by computing the contribution of each feature to the prediction. The SHAP explanation method computes Shapley values using coalitional game theory.

The aim is to fairly determine the effect of every single player on the overall team result. It is assumed that \( n \) players play a cooperative game. The outcome of the game is referred to as \( V(D) \), where \( \{D = 1, \cdots, n\} \) denotes the aggregated set of players. \( \Phi \) is the contributed value of each player to the outcome of the game. An intuitive method is the leave-one out (LOO) method (Cook, 1977), in which the game is first played with all players, and then with the entire set of players but without the player at interest \( i \).

It can be seen from equation (1) that the value of each player is the difference between the game result with the entire dataset minus the game result without the player at interest.

\[
\Phi_i = V(D) - V(D \setminus \{i\}) \tag{1}
\]

To fairly distribute the values of all players, the sum of all individual values \( \Phi_i \) is required to correspond to the overall result of the team, which can be seen in equation (2). The LOO method does not meet this criterion.

\[
V(D) = \sum_{i=1}^{n} \Phi_i \tag{2}
\]

The Shapley values offer an alternative approach, which fulfills this criterion. To fairly distribute the values of the players, each Shapley value considers all subsets \( S \) of players. The weighted sum of the individual performances in the subsets then gives the player’s overall individual performance. Shapley values are thus defined according to equation (3).
The feature values of a data instance act as players in a coalition. Shapley values tell us how to distribute the “payout” (= the prediction) among the features fairly. A player can be an individual feature value, e.g., for tabular data. A player can also be a group of feature values. For example, to explain an image, pixels can be grouped into superpixels and the prediction distributed among them. Using the Python SHAP package\(^1\), it is possible to visualize feature attributions such as Shapley values as \textit{forces}. Each feature value is a \textit{force} that either increases or decreases the prediction. The prediction starts from the baseline. The baseline for Shapley values is the average of all predictions. We use SHAP to explain individual predictions, following Rodríguez-Pérez and Bajorath (2020).

Difficulties in interpreting machine learning (ML) models and their predictions limit the practical applicability of and confidence in ML in various fields, especially in reward crowdfunding literature (see Table 1). The SHAP approach enables the identification and prioritization of features that determine compound classification and activity prediction using any ML model (Rodríguez-Pérez & Bajorath, 2020; Vega García & Aznarte, 2020). This is achieved by explaining the model’s outcome using the concept of additive feature attribution (Cha et al., 2021). Machine learning algorithms extract information from data, and a major concern is understanding how a specific model operates, i.e., how the model makes the decision.

3. Method

3.1. Data

The data collection for this work included the Brazilian reward crowdfunding platform Catarse, which is the largest crowdfunding community in the country since the

\[ \Phi_i = \sum_{S \subseteq D \setminus \{i\}} \frac{V(S \cup \{i\}) - V(S)}{\binom{n-1}{|S|}} \]
beginning of its operations, in 2011, it has raised more than R$ 136 million (approximately USD 43.8 million) through the financing of approximately 13,500 campaigns. The fundraising system on this platform is of the AoN (All-or-Nothing) type (Hemer, 2011). That is, if the financial target $TA$ established by the entrepreneurs, is not reached within the stipulated period, the campaign is canceled and the pledgers receive their resources back or are given credit to finance other campaigns. On the other hand, entrepreneurs of unsuccessful campaigns do not receive any financial resources. We work with three main types of independent variables: i) campaign attributes, ii) geographic localization attributes (see the geographic localization of the reward crowdfunding campaigns is shown in Figure 1), and iii) Market sentiment. In total, we considered 4,193 campaigns, based in 415 Brazilian cities. These campaigns managed to raise more than R$ 38 million (approximately USD 12.3 million) between 2011 and 2015.

3.2. Variables

The variables used in this study were selected considering the literature on model classes to predict the success of reward crowdfunding campaigns. Our three groups of explanatory variables to compose our predictive model of Campaign success$_i$ is as following. The first group presents the attributes of the campaigns: # of Pledges$_i$, Goal$_i$, Rewards$_i$, Category$_i$, and Duration$_i$. The second group of variables concerns geographic aspects that characterize the campaign’s localization: GDP per Capita$_i$, Gini$_i$, Elderly population$_i$, % of Illiteracy$_i$, Population$_i$, and City area$_i$. The third group of variables refers to the Sentiment of the pledger, measured from: mainstream (MMNS) and social media (SMNS) at Day $d$, when campaign $i$ is launched.

Campaign success$_i$ is the dependent variable that informs the success (treated as the achievement of the financial goal $TA$) of a campaign. It is a dummy variable and we assign 1 to it if the amount collected was equal to or greater than the target amount.
(i.e. the financial goal) $T_A$, and 0 otherwise (Strausz, 2017).

$Goal_i$ is the ln of the amount (in R$) targeted for the financing of the $i$-th campaign. This variable was chosen to compose the model because it is believed that the goal is an element that can influence the success of the collective collection of financial resources. (Giudici, Guerini, & Rossi-Lamastra, 2018; G. Li & Wang, 2019; Mollick, 2014).

$\# of Pledges_i$ is the ln of the number of supports received by the campaigns at the end of the fundraising campaign. It was selected considering that it can be considered an element that attracts contributions and reduces uncertainty about the fundraising process by crowdfunding (Colombo, Franzoni, & Rossi-Lamastra, 2015; Josefy, Dean, Albert, & Fitza, 2017).

$Rewards_i$ is the number of rewards offered during the $i$ campaign fundraising period. According to Frydrych, Bock, Kinder, and Koeck (2014), material rewards (or not) can serve to attract more pledgers to fund campaigns.

$Category_i$ indicates the classification of the campaign according to the classification of the financing campaigns managed by the Catarse platform. Thus, the campaigns were grouped according to the following categories: music (814), cinema and video (753), theater (388), literature (342), comics (272), community (222), art (215), photography (109), games (109), dance (61), circus (12), architecture and urbanism (34), carnival (39), science and technology (80), design (60), education (146), sport (60), events (152), gastronomy (7), journalism (103), environment (66), mobility and transport (26), fashion (26) and social businesses (97).

$Duration_i$ is the number of days the campaign takes to reach the target amount $T_A$.

$GDP per Capita_i$ is the ln of the wealth (in R$) produced in the individual’s city that contributes financially to a crowdfunding campaign. We followed the arguments of Mollick (2014), which suggested that wealthier areas may participate more in funding by crowdfunding. Accordingly we included this variable because we believe that individuals residing in wealthier regions can make greater contributions to (Mendes-Da-Silva et al., 2016), in addition we think that regions with different levels of income should have different dynamics for success or failure for crowdfunding.
Gini\(_i\) is the proxy for the level of concentration of household income per capita in the cities that are the headquarters of the campaigns that seek collective financing. This variable was extracted from the last census conducted by the Brazilian Institute of Geography and Statistics (IBGE). According to Mollick (2014) and Burtch et al. (2014), information concerning income and spatial location can help to understand the disproportionate concentration of collective enterprises and should reveal important economic information about the dynamics of crowdfunding.

**Elderly population\(_i\)** is the percentage of the elderly population in each city where crowdfunding campaigns were developed. This variable was considered because the rewards model can be influenced by demographic variables, such as the age group (Gamble, Brennan, & McAdam, 2017). In general, entrepreneurs are young and with limited availability of capital (Gamble et al., 2017). These entrepreneurs are likely to count on financial support from people with some financial independence in their financing campaigns. In this way, it makes sense to think that older people who have some kind of income can contribute to the projects, especially if they maintain a family relationship with the entrepreneur (Agrawal, Catalini, & Goldfarb, 2015).

**% of Illiteracy\(_i\)** is the percentage of illiteracy in the city of origin of the crowdfunding enterprise. This variable was selected in the modeling because, according to Florida (2002), human capital must have a minimum level of education to improve its importance in society and the economy of a given region. In other words, it is reasonable to think that the level of education of individuals in a city can influence their participation in crowdfunding, whether as an entrepreneur or investor.

**Population\(_i\)** is the ln of the estimated population of the city of origin of the campaign (Giudici, Guerini, & Rossi Lamastra, 2013; Giudici et al., 2018), according to official data from IBGE census in 2015. **City area\(_i\)** is the ln the geographic area in \((km^2)\) of the city of origin of the campaign (Mendes-Da-Silva et al., 2016).

We use two different proxies for net sentiment of the pledger’s textual sentiment, measured on the day \(d\) on which the \(i\) campaign is launched on the crowdfunding platform, by using Natural Language Processing-NLP (Y. Li, Thomas, & Liu, 2020): i) Mainstream Media Net Sentiment \((MMNS_d)\), calculated based on news captured on mainstream media, and ii) social media Net Sentiment \((SMNS_d)\) based on social
media posts. We used the feeling of the backer (pledger) on day \( d \), when the campaign was launched, measured through textual analysis of traditional and social media news content on a daily basis relying on the arguments of X. Yuan, Wang, Yin, and Wang (2021) arguments. This proxy captures sentiment and confidence in the national context through news with a positive or negative tone. In this sense, we built two classes of metrics: an indicator of backer sentiment on the day the campaign was launched via traditional media news of wide reach, by the largest newspaper in circulation in Brazil, and a measure based on one of the largest social media platforms, Twitter. We used the procedure adopted by Loughran and McDonald (2011), which in accordance with Kearney and Liu (2014) presents advances in relation to the pioneering work of Tetlock (2007), which proposed the method for carrying out the analysis of the speech content, based on the frequency of words and assigning a mathematical weight \( W_{j,k} \) to the terms found in text content, according to equation (4).

\[
W_{j,k} = \begin{cases} 
1 + \log(tf_{j,k}) \\
1 + \log(a_k) \\
0 & \text{if } tf_{j,k} \geq 1 \\
0 & \text{otherwise}
\end{cases}
\]  

(4)

where: \( W_{j,k} \) = weight of the term of a word \( j \) in a document \( k \); \( tf_{j,k} \) = total events of a \( j \) word in a \( k \) document; \( a_k \) = proportion of words counted in a \( k \) document; \( N \) = total documents in the sample; \( df_j \) = total documents with at least of one occurrence of the word \( j \). Thus, we measure the pledger of reward crowdfunding campaigns via social and traditional media, based on the largest newspaper published in Brazil, \( O Estado de São Paulo (\@Estadao) \), and on the activity of access and filing in the directory of the Twitter Streaming API service, from more than 101 thousand tweets published on the campaign launch dates. We accessed the official Twitter account of this newspaper, which disseminates and shares news, texts and opinions on economics, politics, culture and society, for the elaboration of our social media dataset. With this procedure, we obtained two proxies for the net effect of the textual sentiment, based on the tone of the positive and negative news on the day the campaigns are launched. The 1,054 daily news published in mainstream media used for the pledgers’ sentiment
analysis on the day the campaigns were launched, were manually collected from the
cover of the newspaper *O Estado de São Paulo* (Nascimento, Rodrigues, & Kraemer,
2015). Founded in 1870, this newspaper has an average daily circulation of 300,000
copies (Estadao, 2021), is available in electronic format on the estadao.com domain
and is located in the financial center of Latin America. This newspaper is part of a
private news agency (*Grupo Estado*), with extensive capacity to replicate information
of a varied nature throughout the Brazilian territory, dealing with economic, political,
social and cultural issues.

### 3.3. Experimental Methodology

Table 2 shows the number of successful and unsuccessful campaigns per year. Despite
the increase in the number of campaigns over the years (2011 had 298 campaigns,
while 2015 had 1404), the classes are balanced, i.e., the two classes (success and un-
success) have a similar number of instances. The whole dataset contains a total of
4,193 campaigns.

<table>
<thead>
<tr>
<th>TABLE 2 ABOUT HERE</th>
</tr>
</thead>
</table>

The experimental evaluation is conducted using four different scenarios that aim to
evaluate the behavior of the models under different sizes for the training and testing
datasets. In the first scenario (I), all the campaigns of 2011 are used to train the
models, and the campaigns from 2012 to 2015 are used for evaluating the models. For
the second scenario (II), the training dataset is from the years 2011 and 2012, and the
testing dataset is composed of the years from 2013 to 2015. The other two scenarios
follow the same rationale, and in the last scenario, all years are used to train the
models except 2015 which is then used to evaluate the machine learning algorithms.
Table 3 shows the number of campaigns for each one of the four scenarios.

<table>
<thead>
<tr>
<th>TABLE 3 ABOUT HERE</th>
</tr>
</thead>
</table>

14
We considered the following learning algorithms in this study: k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Multilayer Perceptron Neural Network (MLP), SGD, Decision Trees (DT), AdaBoost (ADA), Random Forest (RF), and XGBoost (XGB). We also evaluated multiple classifiers systems, such as Single Best (SB) (R. M. Cruz et al., 2018), Static Selection (StaticS) (R. M. Cruz et al., 2018), Overall Local Accuracy (OLA) (Woods et al., 1997), Local Class Accuracy (LCA) (Woods et al., 1997), k-Nearest Oracle Union (KNORAU) (Ko et al., 2008), k-Nearest Oracle-Eliminate (KNORAE) (Ko et al., 2008), and MetaDES (R. M. O. Cruz et al., 2015). These models were selected since they belong to different families of learning machines, and they are among the best general purpose classification techniques as reported in Fernández-Delgado et al. (2014). Multiple classifier systems, in particular, dynamic selection techniques (R. M. Cruz et al., 2018) have been shown to outperform single classifiers (Kittler, Hatef, Duin, & Matas, 1998; L. I. Kuncheva, 2014; Woźniak, Graña, & Corchado, 2014b). Moreover, to the best of our knowledge, dynamic selection techniques are evaluated herein for the first time to predict campaign success. We used the implementation provided by scikit-learn (Pedregosa et al., 2011) and the DESlib library (R. M. O. Cruz, Hafemann, Sabourin, & Cavalcanti, 2020).

We used a grid search to find the best set of hyperparameters for each machine learning model. Table 4 shows the hyperparameter values considered for each classifier model. For each model, the five fold cross-validation procedure was performed on the training set, and the hyperparameter configuration that obtained the best performance was selected.

All the experiments were evaluated using three measures: accuracy, area under the ROC Curve (AUC), and F-score (F1). Accuracy is a commonly used measure for classification tasks, and it is defined as the ratio between the number of correctly
classified instances and the total number of evaluated instances. The ROC curve is a graph that shows the trade-off between the true positive rate and the false-positive rate for different classification thresholds. In contrast to the ROC curve that provides a 2-dimensional plot, AUC provides a model’s performance measure over all possible classification thresholds. F1 is calculated as a combination of precision and recall.

4. Results

The prediction results can be classified into four categories: i) true positive, when an occurrence is correctly predicted; ii) true negative, when a nonoccurrence is correctly predicted; iii) false-positive, when a nonoccurrence is incorrectly predicted; and iv) false negative, when an occurrence is incorrectly predicted. For the test dataset, the performance of the optimized ML models was evaluated based on three measures: accuracy (ACC), area Under the ROC curve (AUC), and F-score (F1).

Optimized ML models were evaluated using the test dataset and the results are shown in Table 5. Independent of the scenario and used model, all the results show a strong performance for all evaluation metrics (i.e., ACC, AUC, and F1). The best results per scenario are highlighted in bold, and the second-best results are underlined. Among the single classifiers, SVM obtained good accuracy for all scenarios, especially for the AUC metric. Overall, all the machine learning algorithms achieve very competitive rates, except the k-nearest Neighbor (kNN). RF, XGB, and AdaBoost obtained better results than DT, in general. Thus, given that they use DT as base classifiers, we can infer that these algorithms benefit from using many DTs to predict the instance’s label.

| TABLE 5 ABOUT HERE |

Compared to the single classifiers, the multiple classifier systems deserve special attention since they received, in general, the best results. In particular, for the ACC and F1 measures, the best results for all four scenarios were reached by combining the
classifiers. Among the combination strategies, Meta-DES was the technique with the best overall result.

Scenario IV (Table 3) is a more realistic scenario since it predicts the outcomes for the next year using all the past data, i.e., the training process uses 2,789 campaigns (from 2011 to 2014) to predict the outcomes of 1,404 campaigns from 2015. KNORAU obtained the best rates for accuracy and F-score, 90.81 and 90.79, respectively, followed by the Meta-DES. More precisely for the accuracy measure, Meta-DES correctly classified 1,273 campaigns and 131 campaigns had their labels predicted incorrectly. The errors are distributed as follows: 112 successful campaigns were classified as unsuccessful, and 19 unsuccessful were classified as successful.

4.1. Discussion

In this section, we analyze how the model’s outcome and predictions relate to the variables used to train the 15 different ML models. This analysis is conducted based on the computation of the Shapley values. Consequently, we use the SHAP library\(^2\) in Scenario IV (see Table 3), and the selected model was the Meta-DES (see Table 4) since it obtained the best results (see Table 5). SHAP does a great job in decoding the strength of the influence of the input variables in the predictions. SHAP values determine the feature importance of a feature by comparing what a model predicts with and without the feature. We now explain how SHAP can help us to obtain a local knowledge, i.e., how the likelihood of obtaining a higher/lower reward crowdfunding success of each observation is formulated. Explanations obtained by the deep SHAP method are represented graphically.

Figure 2a shows, for a certain successful campaign, the force SHAP plot for one test instance extracted from scenario IV using the Meta-DES model. This plot shows how each feature contributes to the model prediction having the mean model prediction over the training dataset as a baseline. Likewise, Figure 2b shows the force SHAP plot for one test instance extracted from scenario IV using the Meta-DES model, but for a failed campaign. These plots show how each feature contributes to the model prediction with the mean model prediction over the training dataset as a baseline.

\(^2\)https://github.com/slundberg/shap
For the campaign represented by Figure 2b, the base value (which is considered to be the mean model prediction over the training dataset) is 0.4858, and the Meta-DES model prediction for the test instance, \( f(x) \), is 0.10. Thus, red variables push the result higher than the “base value”, while blue variables push the result lower. In other words, the blue variables, such as \# of Pledges (0.296) and GDP per capita (0.9269), are associated with the attribution of negative SHAP values, i.e., the Campaign success, is expected to decrease consequently. In contrast, the red variable, MMNS (−1.155), Population (0.9338), and campaigns started in March (3.066) are associated with the attribution of positive SHAP values, i.e., the Campaign success, is expected to increase.

FIGURE 2 ABOUT HERE

In contrast to Figure 2a and Figure 2b that show the prediction of a single test instance, Figure 3 shows the whole test set, where the 1,404 test instances are sorted by the explanation similarity. When the \# of Pledges are low, the model tends to indicate that the campaign will not be a success. However, when the \# of Pledges are high (represented in red at the middle of the plot), the model tends to predict that the campaign will be a success.

FIGURE 3 ABOUT HERE

Figure 4 shows the SHAP summary plot that combines feature importance with feature effects with a global view of the instances. Thus, in this Figure, each point is a Shapley value for a feature and an instance. The position on the y-axis is determined by the feature and the position on the x-axis is determined by the Shapley value. The color represents the value of the feature from low to high. Overlapping points are spread in the y-axis direction and as result, we obtain a sense of the distribution of the Shapley values per feature. The features are ordered according to their importance. We
are able to see the importance of the features using all the points in the training data, and the variables are ranked in descending order. Thus, \(# \text{ of Pledges} \) and \( \text{Goal}_i \) are the most relevant attributes to predict a campaign’s success. The horizontal spread per feature shows the direction of the impact on the model output. For instance, higher values of \(# \text{ of Pledges} \) (red) are more strongly associated with \( \text{Campaign success} \). On the other hand, the \( \text{Goal}_i \) feature is inversely proportional to the success of a campaign, i.e., a low \( \text{Goal}_i \) likely leads to a successful campaign. This finding is compatible with previous intuitions and scientific knowledge, which tells us that the larger a pledge, the more prone to success a crowdfunding campaign is (Mendes-Da-Silva et al., 2016).

\[ \text{FIGURE 4 ABOUT HERE} \]

\[ \text{FIGURE 5 ABOUT HERE} \]

5. Concluding remarks

Concerns about reward crowdfunding campaign success and its drivers have risen in recent years in developed and emerging countries. Several studies address the crowdfunding campaign success forecasting problem, using techniques that range from traditional statistical approaches to more recent ones such as machine learning models, which have proven useful in this framework (see Table 1). However, there is still a gap
between the information provided by these models and the needs of the stakeholders. Not only is an accurate model needed but also a model that provides interpretable results on reward crowdfunding campaigns success and their drivers. Concerning the interpretation of the prediction of reward crowdfunding campaign success, results show that the \# of Pledges has the greatest impact on the predictions, being directly proportional to reward crowdfunding campaigns success (see Figure 4). This is aligned with the current understanding of the phenomena under study.

Among the fifteen machine learning models evaluated, Meta-DES deserves special attention since it achieved the best overall results. Meta-DES belongs to a class of machine learning methods that uses multiple models and selects the best ones per test instance; these selected models are combined to predict the class of the campaign under evaluation. Therefore, compared to traditional single models, such as multilayer perceptrons, decision trees, k-nearest neighbor, random forest, and others, the models that use multiple classifiers present as a valuable alternative to predict the campaign success, in particular, those that dynamically select the most competent models per campaign.

Finally, we find a light explanatory power in textual sentiment, both in terms of mainstream media and social media. As we have not found research discussing the effects of mass media on the success of the reward crowdfunding campaign, we understand that future research around this topic can make contributions to machine learning and crowdfunding literature. This proposal can also be interesting for institutions and organizations concerned with funding for new businesses, especially in emerging economies, where institutional environment is a significant drawback for entrepreneurs.

References


Ba, Z., Zhao, Y. C., Song, S., & Zhu, Q. (2021). Understanding the determinants of online medical crowdfunding project success in china. Information Processing & Management,


Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Édouard


28
Table 1.: Literature related to reward crowdfunding campaign success forecasting and methods. Methods: SA = Survival Analysis, OLS = Ordinary Least Square, L = Logit, P = Probit, DA = Discriminant Analysis, T = Tobit, HMR = Hierarchical Multiple Regression, 2SLS = Two-Stage Least Squares, ML = Machine Learning. Countries: USA = United States of America, GER = Germany, CH = China, FIN = Finland.

<table>
<thead>
<tr>
<th>Authorship</th>
<th>Platforms (country)</th>
<th>Method</th>
<th>AoN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burtch, Ghose, and Wattal (2013)</td>
<td>Journalism (USA)</td>
<td>2SLS</td>
<td>no</td>
</tr>
<tr>
<td>Crosetto and Regner (2014)</td>
<td>Startnext (GER)</td>
<td>P</td>
<td>yes</td>
</tr>
<tr>
<td>Mollick (2014)</td>
<td>Kickstarter (USA)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>Colombo et al. (2015)</td>
<td>Kickstarter (GER)</td>
<td>T and P</td>
<td>yes</td>
</tr>
<tr>
<td>Hörisch (2015)</td>
<td>Indiegogo (USA)</td>
<td>L</td>
<td>no</td>
</tr>
<tr>
<td>Zvilichovsky, Inbar, and Barzilay (2015)</td>
<td>Kickstarter (USA)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>Hobbs, Grigore, and Molesworth (2016)</td>
<td>Kickstarter (USA)</td>
<td>DA</td>
<td>yes</td>
</tr>
<tr>
<td>Mendes-Da-Silva et al. (2016)</td>
<td>Catarse (Brazil)</td>
<td>OLS</td>
<td>yes</td>
</tr>
<tr>
<td>Shi and Guan (2016)</td>
<td>Jing Dong (CH)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>H. Yuan et al. (2016)</td>
<td>Dreamore (CH)</td>
<td>ML</td>
<td>yes</td>
</tr>
<tr>
<td>Courtney, Dutta, and Li (2017)</td>
<td>Kickstarter (USA)</td>
<td>L and P</td>
<td>yes</td>
</tr>
<tr>
<td>Skirnevskyi, Bendig, and Bretted (2017)</td>
<td>Kickstarter (USA)</td>
<td>L and T</td>
<td>yes</td>
</tr>
<tr>
<td>Bi, Liu, and Usman (2017)</td>
<td>zhongchou.com (CH)</td>
<td>HMR</td>
<td>no</td>
</tr>
<tr>
<td>Chan, Park, Patel, and Gomulya (2018)</td>
<td>Kickstarter (USA)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>Clauss et al. (2018)</td>
<td>Visionbakery (GER)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>Crosetto and Regner (2018)</td>
<td>Startnext (GER)</td>
<td>P</td>
<td>yes</td>
</tr>
<tr>
<td>de Larrea, Altin, and Singh (2019)</td>
<td>Kickstarter (USA)</td>
<td>HMR</td>
<td>yes</td>
</tr>
<tr>
<td>Lagazio and Querci (2018)</td>
<td>Indiegogo (USA)</td>
<td>P</td>
<td>no</td>
</tr>
<tr>
<td>Oo, Allison, Sahaym, and Juasrikul (2019)</td>
<td>Kickstarter (USA)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>Da Cruz (2018)</td>
<td>Kickstarter (USA)</td>
<td>P</td>
<td>yes</td>
</tr>
<tr>
<td>N. Wang, Li, Liang, Ye, and Ge (2018)</td>
<td>Dreamore (CH)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>T.-L. Yeh, Chen, and Lee (2019)</td>
<td>Makuake (Japan)</td>
<td>L</td>
<td>yes</td>
</tr>
<tr>
<td>J.-Y. Yeh and Chen (2020)</td>
<td>Indiegogo (USA)</td>
<td>ML</td>
<td>yes</td>
</tr>
<tr>
<td>Ba, Zhao, Song, and Zhu (2021)</td>
<td>Tencent GongYi (CH)</td>
<td>OLS</td>
<td>yes</td>
</tr>
<tr>
<td>Shin et al. (2021)</td>
<td>Mesenaatti (FIN)</td>
<td>OLS</td>
<td>no</td>
</tr>
<tr>
<td>Tafesse (2021)</td>
<td>Kickstarter (USA)</td>
<td>NBR</td>
<td>yes</td>
</tr>
<tr>
<td>Felipe et al. (2022)</td>
<td>Catarse (Brazil)</td>
<td>SA</td>
<td>yes</td>
</tr>
<tr>
<td>This study</td>
<td>Catarse (Brazil)</td>
<td>ML</td>
<td>yes</td>
</tr>
</tbody>
</table>
Table 2.: Number of campaigns (instances) per year. This table shows the number of successful and unsuccessful campaigns per year. Despite the increase in the number of campaigns over the years (2011 had 298 campaigns, while 2015 had 1,404), the classes are balanced, i.e., the two classes (success and unsuccess) have a similar number of instances. The whole dataset contains a total of 4,193 campaigns.

<table>
<thead>
<tr>
<th>Year</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of campaigns</td>
<td>298</td>
<td>531</td>
<td>762</td>
<td>1198</td>
<td>1404</td>
</tr>
<tr>
<td>Success</td>
<td>156</td>
<td>294</td>
<td>464</td>
<td>650</td>
<td>652</td>
</tr>
<tr>
<td>Unsuccess</td>
<td>142</td>
<td>237</td>
<td>298</td>
<td>548</td>
<td>752</td>
</tr>
</tbody>
</table>
Table 3.: **Four different scenarios for dividing the dataset into training and testing.** This table shows the number of campaigns for each one of the four scenarios. In the first scenario (I), all the campaigns that occurred in 2011 are used to train the models, and the campaigns from 2012 to 2015 are used for evaluating the models. In the second scenario (II), the training dataset is composed of the campaigns from the years 2011 and 2012, and the testing dataset is composed of campaigns from 2013 to 2015. The other two scenarios follow the same rationale, and in the last scenario, all years are used to train the models except 2015 which is used to evaluate the machine learning algorithms.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Years (number of campaigns)</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2011 (298)</td>
<td>2012-2015 (3895)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>2011-2012 (829)</td>
<td>2013-2015 (3363)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>2011-2014 (2789)</td>
<td>2015 (1404)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1.: Geographical distribution of reward crowdfunding campaigns in Brazil. Source: Prepared by the authors, based on data obtained from the Catarse platform, using the ArcGIS ArcMap 10.0 tool (Zieg & Zawada, 2021). Note: This figure shows the spatial dispersion of 4,193 crowdfunding campaigns on the Catarse platform in the period 2011-2015. See also Table 1).
Table 4.: The hyperparameter grid considered for the experiments. Table 4 presents the hyperparameter values considered for each classifier model. For each model, the fivefold cross-validation procedure was performed on the training set, and the hyperparameter configuration that obtained the best performance was selected. KNN = k-Nearest Neighbors, SVM = Support Vector Machines, MLP = Multilayer Perceptron Neural Network, SGD = Stochastic Gradient Descent, DT = Decision Trees, ADA = AdaBoost, RF = Random Forest, XGB = XGBoost, SB = Single Best, StaticS = Static Selection, OLA = Overall Local Accuracy, LCA = Local Class Accuracy, KNORAU = k-Nearest Oracle Union, KNORAE = k-Nearest Oracle-Eliminate, Meta-DES = Meta Learning Dynamic Ensemble Selection.

<table>
<thead>
<tr>
<th>Method</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>criterion: [gini, entropy] splitter: [best, random] max-depth: [3, 4, 5, 6, 8, 10, 12, 15] and none</td>
</tr>
<tr>
<td>RF</td>
<td># estimators: [100, 200, …, 500] max-depth = max-depth: [3, 4, 5, 6, 8, 10, 12, 15] and none</td>
</tr>
<tr>
<td>XGB</td>
<td># booster: decision tree learning-rate: [0.05, 0.10, 0.15, 0.20, 0.25, 0.30] max-depth: [3, 4, 5, 6, 8, 10, 12, 15]</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>base-estimator: decision tree learning-rate: [0.05, 0.10, 0.15, 0.20, 0.25, 0.30] max-depth: [3, 4, 5, 6, 8, 10, 12, 15] # estimators: [10, 50, 250, 1000]</td>
</tr>
<tr>
<td>MLP</td>
<td>hidden-layer-sizes: [50, 100, 200] activation function: [identity, logistic, tanh, relu] solver: [lbfgs, sgd, adam] alpha: [0.0001, 0.001, 0.01, 0.05, 0.5, 1.0] learning-rate: [constant, invscaling, adaptive]</td>
</tr>
<tr>
<td>SVM</td>
<td>kernel: [poly, sigmoid, rbf] gamma: [1, 1.0, 1.0, 0.01, 0.001] c: [0.1, 1, 10, 100]</td>
</tr>
<tr>
<td>SGD</td>
<td>penalty: [L2, LI, elasticnet] alpha: [0.0001, 0.001, 0.00001, 0.01] loss: [squared_epsilon_insensitive, hinge, log, modified_huber, squared_hinge, perceptron]</td>
</tr>
<tr>
<td>KNN</td>
<td>n_neighbors: [1, 3, 5, 7, 9, 13] weights: [uniform, distance]</td>
</tr>
<tr>
<td>Static Selection, Single Best</td>
<td># estimators: [10, 20, 30, 50, 70, 100] base_estimators: [DecisionTree, Perceptron, LogisticRegression, GaussianNB]</td>
</tr>
<tr>
<td>OLA, LCA, KNORAU, KNORAE, Meta-DES</td>
<td>k: [3, 7, 5, 9, 13] dfp: [True, False] # estimators: [10, 20, 30, 50, 70, 100] base_estimators: [DecisionTree, Perceptron, LogisticRegression, GaussianNB]</td>
</tr>
</tbody>
</table>
Table 5.: **Classifiers’ results per scenario.** This table presents the accuracy (ACC), area under the receiver operating characteristic-ROC curve (AUC), and F1-score (F1) per scenario for each model. The best and second best results per scenario are in bold and underlined respectively. KNN = k-Nearest Neighbors, SVM = Support Vector Machines, MLP = Multilayer Perceptron Neural Network, SGD = Stochastic Gradient Descent, DT = Decision Trees, ADA = AdaBoost, RF = Random Forest, XGB = XGBoost, SB = Single Best, StaticS = Static Selection, OLA = Overall Local Accuracy, LCA = Local Class Accuracy, KNORAU = k-Nearest Oracle Union, KNORAE = k-Nearest Oracle-Eliminate, Meta-DES = Meta Learning Dynamic Ensemble Selection. For information regarding scenarios I, II, III and IV, see Table 3.

<table>
<thead>
<tr>
<th>Models</th>
<th>ACC</th>
<th>AUC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>1. KNN</td>
<td>73.76</td>
<td>73.45</td>
<td>68.13</td>
</tr>
<tr>
<td>2. SVM</td>
<td>90.91</td>
<td>88.49</td>
<td>88.39</td>
</tr>
<tr>
<td>3. MLP</td>
<td>89.83</td>
<td>89.29</td>
<td>90.19</td>
</tr>
<tr>
<td>4. SGD</td>
<td>87.44</td>
<td>88.28</td>
<td>89.39</td>
</tr>
<tr>
<td>5. DT</td>
<td>88.52</td>
<td>89.89</td>
<td>88.97</td>
</tr>
<tr>
<td>6. ADA</td>
<td>89.21</td>
<td>90.54</td>
<td>90.08</td>
</tr>
<tr>
<td>7. RF</td>
<td>86.54</td>
<td>87.78</td>
<td>87.89</td>
</tr>
<tr>
<td>8. XGB</td>
<td>89.57</td>
<td>91.05</td>
<td>90.69</td>
</tr>
<tr>
<td>9. SB</td>
<td>90.93</td>
<td>90.60</td>
<td>90.04</td>
</tr>
<tr>
<td>10. StaticS</td>
<td>90.50</td>
<td>91.11</td>
<td>90.31</td>
</tr>
<tr>
<td>11. OLA</td>
<td>90.65</td>
<td>91.08</td>
<td>90.04</td>
</tr>
<tr>
<td>12. LCA</td>
<td>89.34</td>
<td>90.33</td>
<td>89.35</td>
</tr>
<tr>
<td>13. KNORAU</td>
<td>90.57</td>
<td>91.46</td>
<td>90.62</td>
</tr>
<tr>
<td>14. KNORAE</td>
<td>89.80</td>
<td>91.31</td>
<td>90.54</td>
</tr>
<tr>
<td>15. Meta-DES</td>
<td>90.60</td>
<td>91.61</td>
<td>90.81</td>
</tr>
</tbody>
</table>
(a) Explanation of test prediction for a certain successful reward crowdfunding campaign.

(b) Explanation of test prediction for a certain failed reward crowdfunding campaign.

Figure 2.: **Explanation of two test predictions for the scenario IV generated by the Meta-DES model.** This figure shows the force SHAP plot for two test instances (i.e., reward crowdfunding campaigns) extracted from scenario IV (see Table 3) using the Meta-DES model. The associated Shapley value of a feature is visualized using the length of an arrow. Feature expressions with large Shapley values have strong effects on the individual prediction and are shown in the middle of SHAP force plots. Figure 2a: SHAP force plots show that the individual prediction, via Meta-DES prediction, \( f(x) = 1.00 \) consists of the sum of all feature Shapley values and the average model prediction (base value = 0.4858). Figure 2b: SHAP force plots show that the individual prediction, via Meta-DES prediction, \( f(x) = 0.10 \) consists of the sum of all feature Shapley values and the average model prediction (base value = 0.4858). Thus, for both Figure 2a and Figure 2b, red variables push the result higher than the “base value”, while blue variables push the result lower.
Figure 3.: **Explanation of the whole test dataset predictions for scenario IV generated by the Meta-DES model.** This is stacked SHAP explanations clustered by explanation similarity. Each position on the x-axis is an instance of the data, and the y-axis is the model output value. Red SHAP values increase the prediction, and blue values decrease it. This figure shows the whole test set, where the 1,404 test instances are sorted by the explanation similarity. When the # of Pledges are low, the model tends to indicate that the campaign will not be a success. However, when the # of Pledges are high (represented in red at the middle of the plot), the model tends to predict that the campaign will be a success. Three cluster stands out. On the left and right are two groups with a low predicted campaign success.
Figure 4: **SHAP summary plot for the Meta-DES model.** This figure visualizes Shapley values as the success of the reward crowdfunding campaign and is (one dot per campaign) dependent on feature values for a subset of features of the Meta-DES model applied to scenario IV (Table 3). This figure shows the SHAP summary plot that combines feature importance with feature effects, and shows a global view of the instances. Thus, in this figure, each point visualizes a Shapley value for one subject and one feature. The position on the $y$-axis is determined by the feature and on the $x$-axis by the Shapley value. The color represents the value of the feature from low to high. In other words, the color of the points depends on the feature values, and the $x$-axis shows the calculated Shapley values. The $y$-axis represents both the features, ordered by the mean absolute Shapley values and their distribution. *Comic cartoons, games, community, theater, music* are the campaign categories defined by the entrepreneur based on a typology adopted by the crowdfunding platform. We also added a variable to express the day of the week and the month the pledge occurred, with the goal of checking for some kind of calendar effect on the backers’ behavior.
Figure 5.: SHAP feature dependence plot for the predicted outcome of the Meta-DES model, observing two variables, Goal and # of Pledges for scenario IV. This figure shows the marginal effect that the variables, # of Pledges and Goal, have on the predicted outcome of the Meta-DES model. Thus, this plot shows that there is an approximately linear and positive trend between the # of Pledges and the target variable; moreover, it is one can observe that the # of Pledges interacts with Goal frequently.